Gene 494 (2012) 181–189

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Cloning and characterization of homeologous *cellulose synthase catalytic subunit 2* genes from allotetraploid cotton (*Gossypium hirsutum* L.)

Hee Jin Kim [a,*], Barbara A. Triplett [a], Hong-Bin Zhang [b], Mi-Kyung Lee [b], Doug J. Hinchliffe [a,c], Ping Li [a], David D. Fang [a]

[a] USDA-ARS, Southern Regional Research Center, Cotton Fiber Bioscience, 1100 Robert E. Lee Blvd, New Orleans, LA 70124 USA
[b] Department of Soil and Crop Sciences, 2474 TAMU, Texas A&M University, College Station, Texas 77843-2474, USA
[c] USDA-ARS, Southern Regional Research Center, Cotton Chemistry and Utilization, 1100 Robert E. Lee Blvd, New Orleans, LA 70124 USA

ARTICLE INFO

Article history:
Accepted 8 December 2011
Available online 20 December 2011

Keywords:
Cellulose biosynthesis
Cellulose synthase catalytic subunit
Cotton
Gossypium hirsutum
Secondary cell wall

ABSTRACT

Cellulose synthase catalytic subunits (CesAs) are the catalytic sites within a multisubunit complex for cellulose biosynthesis in plants. CesAs have been extensively studied in diploid plants, but are not well characterized in polyploid plants. *Gossypium hirsutum* is an allotetraploid cotton species producing over 90% of the world's cotton fibers. Although *G. hirsutum CesAs* (*GhCesAs*) are responsible for cellulose production in cotton fiber, very limited numbers of *GhCesA* genes have been identified. Here, we report isolating and characterizing a pair of homeologous *CesA2* genes and their full-length cDNAs from allotetraploid cotton. The *GhCesA2*-A_T gene from the A-subgenome and *GhCesA2*-D_T gene from the D-subgenome were screened from a *G. hirsutum* BAC library. These genes shared 92% sequence similarity throughout the entire sequence. The coding sequences were nearly identical, and the deduced amino acid sequences from *GhCesA2*-A_T (1,039 amino acids) and *GhCesA2*-D_T (1,040 amino acids) were identical except four amino acids, whereas the noncoding sequences showed divergence. Sequence analyses showed that all exons of *GhCesA2*-A_T contained consensus splice donor dinucleotides, but one exon in *GhCesA2*-D_T contained nonconsensus splice donor dinucleotides. Although the nonconsensus splice donor dinucleotides were previously suggested to be involved in alternative splice or pseudogenization, our results showed that a majority of *GhCesA2*-A_T and *GhCesA2*-D_T transcripts consisted of functional and full-length transcripts with little evidence for alternative mRNA isoforms in developing cotton fibers. Expression analyses showed that *GhCesA2*-A_T and *GhCesA2*-D_T shared common temporal and spatial expression patterns, and they were highly and preferentially expressed during the cellulose biosynthesis stage in developing cotton fibers. The observations of higher expression levels of both *GhCesA2*-A_T and *GhCesA2*-D_T in developing fibers of one near-isogenic line (NIL) with higher fiber bundle strength over the other NIL with lower fiber bundle strength suggested that the differential expression of genes associated with secondary cell wall cellulose biosynthesis in developing fiber might affect cotton fiber properties.

Published by Elsevier B.V.

1. Introduction

Cellulose, the most abundant biopolymer in nature, organizes into microfibrils in plant cell walls, providing strength and flexibility to plants. Cellulose is the major constituent of paper and textiles, and has become the primary material for producing ethanol from energy crops. Cellulose synthase, a multisubunit enzyme associated with the plasma membrane in plants, plays a pivotal role in cellulose production (Doblin et al., 2002). The catalytic subunits of cellulose synthase (CesAs), originally named CelAs, are central catalysts in the generation of plant cell wall cellulose (Pear et al., 1996; Kumar et al., 2009). *CesA* genes have been extensively studied in diploid plants like *Arabidopsis*, rice and barley (Richmond and Somerville, 2000; Tanaka et al., 2003; Burton et al., 2004). The completion of the *Arabidopsis* genome sequence revealed ten different *CesA* genes in *Arabidopsis* (Richmond and Somerville, 2000). Genetic and biochemical evidence from *Arabidopsis* showed three *CesAs*, *AtCesA1*, *AtCesA3*, and a *AtCesA6*-related *CesA* (either *AtCesA2*, *AtCesA5*, *AtCesA6* or *AtCesA9*) were required for primary cell wall (PCW) cellulose biosynthesis (Burn et al., 2002; Somerville, 2006; Persson et al., 2007), whereas another set of three *CesAs*, *AtCesA4*, *AtCesA7*, and *AtCesA8*, were involved in secondary cell wall (SCW) cellulose biosynthesis in *Arabidopsis* xylem cells (Taylor et al., 2003; Somerville, 2006). In *Populus trichocarpa*, a model tree that is a paleopolyploid that has

Abbreviations: AFIS, Advanced Fiber Information System; BAC, bacterial artificial chromosome; CesA, cellulose synthase catalytic subunit; DPA, days postanthesis; EST, expressed sequence tag; HVI, High-Volume Instrumentation; NIL, near-isogenic line; NMD, nonsense-mediated mRNA decay; PCW, primary cell wall; PTCs, premature termination codons; RT-qPCR, reverse transcription quantitative polymerase chain reaction; SCW, secondary cell wall; TC, tentative consensus; UTR, untranslated region.

* Corresponding author. Tel.: +1 504 286 4276; fax: +1 504 286 4250.
E-mail address: heejin.kim@ars.usda.gov (H.J. Kim).

0378-1119/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.gene.2011.12.018

undergone subsequent diploidization from ancient genome duplication, there are 17 *PtiCesAs* (Kumar et al., 2009). Unlike the extensive studies of *CesAs* in diploid and paleopolyploid plants, a very limited number of *CesA* genes have been identified from polyploid plants that comprise the major agricultural, horticultural, environmental, and bioenergy crops.

Cotton genus (*Gossypium*) consist of five allotetraploid cotton species (2n = 4x = 52) of five species (AD_1 to AD_5) and about forty five diploid species (2n = 2x = 26) belonging to eight different genomes (A, B, C, D, E, F, G, and K) based on chromosome pairing relationships (Wendel and Cronn, 2003). *G. hirsutum* (AD_1 genome), known as upland cotton, accounts for over 90% of the world's cotton production (Zhang et al., 2008). *G. hirsutum* is an allotetraploid, combining an A genome from the maternal diploid parent (possibly *G. herbaceum* A_2 genome) at the time of polyploid formation and a D genome (possibly *G. raimondii* D_5 genome) from the pollen parent (Applequist et al., 2001; Wendel and Cronn, 2003). *Gossypium* species of diploid cotton (A and D genome) and polyploidy cotton (AD genome) have been used for studying plant genome size evolution and plant polyploidization (Adams et al., 2003; Senchina et al., 2003; Hovav et al., 2008; Chaudhary et al., 2009).

Cotton fibers are unicellular trichomes that differentiate from epidermal cells of developing cotton ovules (Kim and Triplett, 2001). Cotton fiber development is divided into four overlapping stages; initiation, PCW biosynthesis for fiber elongation, SCW biosynthesis for cellulose production, and maturation (Naithani et al., 1982). Fiber initiation starts a day before anthesis, and the initials enter into the elongation phase immediately. During the PCW stage, a thin PCW is deposited in elongating fibers and cotton fibers elongate up to 3–6 cm for 2–3 weeks. The SCW stage initiates approximately 14 to 16 days post anthesis (DPA), overlapping the final PCW stage. At the transition from PCW to SCW biosynthesis in cotton fiber, synthesis of other cell wall polymers ceases and the rate of cellulose synthesis in cotton fibers is estimated to increase nearly 100-fold *in vivo* (Meinert and Delmer, 1977). Mature fibers exhibit thickened SCW composed of nearly pure cellulose.

Cotton fibers were suggested to be a model system for studying cellulose biosynthesis (Kim and Triplett, 2001). The first plant *CesA* cDNAs, *GhCesA1* and *GhCesA2,* originally named *CelA1* and *CelA2*, were isolated from developing *G. hirsutum* fibers (Pear et al., 1996). *GhCesA1* (U58283) was a full length cDNA encoding 974 amino acids, and *GhCesA2* (U58284) was a 5′-truncated partial cDNA encoding 685 amino acids. Later, homeologous *GhCesA1* genes in both the A-subgenome and D-subgenome from *G. hirsutum* were isolated from cotton BAC libraries (Grover et al., 2004). However, neither full length *GhCesA2* cDNAs nor the genes *per se* have yet been identified. A very limited number of homeologous genes have been identified from allotetraploid *G. hirsutum* due to technical difficulties in isolating homeologous genes using PCR techniques. PCR based cloning techniques often amplify PCR-induced intergenic chimeras from homeologous genes of allotetraploid *G. hirsutum* cotton due to the high sequence similarity of homeologous genes (Cronn et al., 2002). In addition, the large genomes (approximately 2.5 Gb) of allotetraploid cottons have not been sequenced.

In this study, we screened from a *G. hirsutum* BAC library and identified a *GhCesA2*-A_T gene from the A-subgenome and a *GhCesA2*-D_T gene from the D-subgenome of allotetraploid cotton. Both *GhCesA2*-A_T and *GhCesA2*-D_T were composed of 12 exons and 11 introns. All 11 intron junctions of *GhCesA2*-A_T consisted of the consensus splice donor dinucleotides (GT…AG) required for removing introns from mRNA precursors and ligating exons to form mature mRNA, whereas one of the 11 intron junctions from *GhCesA2*-D_T contained nonconsensus splice dinucleotides (GC…AG) enriched in alternatively spliced genes in human and *C. elegans* (Thanaraj and Clark, 2001; Farrer et al., 2002). Several cotton *CesA* genes containing the nonconsensus GC splice donor dinucleotides were previously classified as pseudogenes due to the absence of the consensus GT splice donor dinucleotides (Cronn et al., 1999). Our results showed that both *GhCesA2*-A_T and *GhCesA2*-D_T were able to produce functional and mature *CesA2* mRNAs in developing cotton fibers regardless of the variation in the splice donor dinucleotides. We also showed that *GhCesA2*-A_T and *GhCesA2*-D_T were detected in both fiber and non-fiber tissues of *G. hirsutum*, and these genes were most expressed in the SCW stage of developing fibers. Furthermore, transcript levels of *GhCesA2*-A_T and *GhCesA2*-D_T were compared between two near-isogenic lines with different fiber bundle strength.

2. Materials and methods

2.1. Plant materials and growth conditions

Upland cotton plants (*G. hirsutum* L. cv. TM-1, MD 52ne, and MD 90ne) were grown in the field at the USDA, ARS, Southern Regional Research Center, New Orleans, LA in 2008 and 2009. Developing bolls were collected by 9 AM at 2 day intervals from 8 through 20 DPA and fibers were immediately harvested and frozen in liquid nitrogen. Fully grown leaves (15 cm in diameter), expanding young leaves (5 cm in diameter), bracts, and petals were also harvested from the field. Cotyledon, hypocotyls and roots were harvested from 1 week old plants grown in a greenhouse at 25 °C to 32 °C. All tissues were frozen in liquid nitrogen, and stored at −80 °C.

2.2. BAC library screening and fingerprinting for homeologous GhCesA2 genes from allotetraploid cotton

GhCesA2 homeologs were screened from a BAC library for *G. hirsutum*, (TM-1) with BamHI-derived inserts in the pCLD04541 (Zhang et al., 2008). The specific overgo (40-mer) and probe sequences for *GhCesA2* were selected using ClustalW and BLAST against the DFCI Cotton Gene Index. The overgo of *GhCesA2* (708-747nt of U58284, 5′-GTCTCTGAGAAACGACCAAAGATGACATGTGATTGCTGGC-3′) was labeled with [^{32}P] dCTP using Klenow at 37 °C for 30 min. The 166 nt of *GhCesA2* (1635–1800 nt of U58284) was used to make radioactive probes by random priming with [^{32}P]-dCTP. After removal of unincorporated nucleotides using a QIA Quick Nucleotide Removal Kit (Qiagen, Valencia, CA), probes were denatured at 95 °C for 10 min and added to a hybridization container. The hybridizations were performed at 50 °C for 18 h in hybridization solution (5X SSC, 0.5% SDS, 25 mM potassium phosphate at pH 6.5, and 5X Denhardt's). Filters were washed with 1X SSC and 0.1% SDS at 50 °C, and then exposed to X-OMAT AR film (Kodak, Rochester, NY). Primary and secondary screenings identified four positive BAC colonies. DNAs from the positive colonies were isolated using a BACMAX™ DNA Purification Kit (Epicentre Biotechnologies, Madison, WI). The isolated BAC DNAs were triple-digested with EcoRI, BamHI, and HaeIII, end-labeled with [^{33}P] dATP using reverse transcriptase for 2 h at 37 °C, and then subjected to 3.5% (w/v) polyacrylamide DNA sequencing gel electrophoresis at 85 W for 100 min. The gel was dried and autoradiographed. The fingerprints of the autoradiographs were scanned into image files using a UMAX Mirage D-16 L scanner and edited using the Image 4.0 (Soderlund et al., 1997). Two different *GhCesA2* genes identified by the fingerprints were sequenced and assembled using SeqMan Pro software (DNASTAR Inc., Madison, WI).

2.3. Verification of GhCesA2 homeologous genes

To locate two different *GhCesA2* genes in allotetraploid cotton subgenomes, the specific sequences for *GhCesA2*-A_T (5′-ACGAGGTACGGCAGACGAGGTTT-3′/5′-AATTCATGTAGATCCAAACTTTC-3′; amplicon, 570 bp) and *GhCesA2*-D_T (5′-TCTGATAATACTG-AACATGGTCGGAGT-3′/5′-TGAATCAGACCCACCCCGTAAAATCTAGTT-3′; amplicon, 470 bp) as well as the conserved sequences among *CesAs* (5′-

TACAGAAGATATCTTAACAGGATTC-3′/5′-GAGAAGTGGTAAAGATGT-GAAGGGG-3′; amplicon, 274 bp) were PCR amplified with genomic DNA templates from allotetraploid *G. hirsutum* TM-1 (AD_1) as well as following diploid species from *G. herbaceum* (A_1), *G. arborium* (A_2), *G. thurberi* (D_1), *G. armourianum* ($D_{2\text{-}1}$), *G. davidsonii* ($D_{3\text{-}d}$), *G. klotzschianum* ($D_{3\text{-}k}$), *G. aridum* ($D_{4\text{-}3}$), *G. raimondii* (D_5), *G. gossypioides* ($D_{6\text{-}1}$), *G. lobatum* ($D_{7\text{-}4}$), *G. trilobum* ($D_{8\text{-}2}$), and *G. laxum* ($D_{9\text{-}5}$).

2.4. Isolation of full length cDNAs of GhCesA2-A_T and GhCesA2-D_T

Total RNA was extracted from 8, 12, 16, and 20 DPA fibers of TM-1 using a Plant Total RNA Kit and DNase I (Sigma, St. Louis, MO). RNA quantity was determined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE). The RNA quality was determined by an RNA integrity number (RIN) (Mueller et al., 2004) using an Agilent Bioanalyzer 2100 with the RNA 6000 Nano 1 Kit Chip (Agilent Technologies Inc., Santa Clara, CA). First strand cDNA was synthesized from total RNAs by priming with oligo dT primer using Thermoscript Reverse Transcriptase (Invitrogene, Carlsbad, CA) at 50 °C. Full-length homeologous cDNAs of *GhCesA2*-A_T *and GhCesA2*-D_T were PCR amplified from the first strand cDNAs with specific primer sets for *GhCesA2*-A_T (5′-TTGGTTTTGCCATGGCTTCAACCAC-CATGG-3′/5′-GAAATTAAATTGAACCAACAAAATCATAGG-3′; amplicon size, 3,264 bp) and *GhCesA2*-D_T (5′-TTGGTTTTGCCATGGCTTCAACCAC-CATGG-3′/5′-TGAATCAGACCCACCCCGTAAAATCTAGTT-3′; amplicon size, 3,313 bp) using AccuPrime™ Taq DNA Polymerase High Fidelity enzyme (Invitrogen, Carlsbad, CA). Both *GhCesA2* cDNAs amplified from 20 DPA fibers were cloned into the pCR-XL-TOPO vector (Invitrogen, Carlsbad, CA) and sequenced.

2.5. Temporal and spatial regulation of GhCesA2-A_T and GhCesA2-D_T

Total RNA extracted from various tissues at different developmental stages of TM-1. Semi-quantitative RT-PCR was performed with specific primer sets for *GhCesA2*-A_T (5′-TCTGATAATACTGAA-CATGGTCGGAGT-3′/5′-GAAATTAAATTGAACCAACAAAATCATAGG-3′; amplicon size, 424 bp) and *GhCesA2*-D_T (5′-TCTGATAATACTGAA-CATGGTCGGAGT-3′/5′-TGAATCAGACCCACCCCGTAAAATCTAGTT-3′; amplicon, 470 bp) that were located at the 12th exon and 3′ UTR.

2.6. Comparison of GhCesA2-A_T and GhCesA2-D_T transcripts between NILs MD 52ne and MD 90ne

Total RNAs were extracted from developing cotton fibers at 2 day intervals from 12 through 18 DPA of both MD 52ne and MD 90ne. For RT-qPCR, first strand cDNA was synthesized using iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA) with random hexamers. The RT-qPCR was performed using iTaq™ SYBR® Green Supermix with ROX (Bio-Rad Laboratories). The specific primer sets were *GhCesA2*-A_T (5′-TGTTGATTTGCTGTGATTCTAAAAGGGATT-3′/5′-GAAATTAAATTGAACCAACAAAATCATAGG-3′; amplicon size, 81 bp) and *GhCesA2*-D_T (5′-TGTTGATTTGCTGTGATTCTAAAAGGGATT-3′/5′-TAGGTGAATCACACATTTTGCTCTTGCATT-3′; amplicon size, 97 bp). Thermal cycler parameters for RT-qPCR were as follows: 95 °C 2 min, 40 cycles of 95 °C 15 s, 60 °C 30 s. A dissociation curve was generated and used to validate that a single amplicon was present for each RT-qPCR reaction. Relative quantification of target gene transcript abundance is expressed as fold-difference and was performed using the comparative Cp method (Bustin et al., 2009). The transcript levels during fiber development between MD 52ne and MD 90ne were normalized with respect to *18S ribosomal RNA* (U42827; 5′-CGTCCCTGCCCTTTGTACA-3′/5′-AACACTTCACCGGACCATTCA-3′; amplicon size, 63 bp). A total of three RT-qPCR reactions were performed at each time point for cotton tissues representing two biological replications and three technical replications. Statistical analyses and construction of graphs were performed using Prism version 3.00 software (GraphPad Software, Inc., San Diego, CA).

2.7. Comparison of cotton fiber properties between G. hirsutum, MD 52ne and MD 90ne

Cotton fiber properties such as bundle strength, fineness, length, and maturity ratio were measured from three biological replications of MD52ne and MD 90ne using a High-Volume Instrumentation (HVI) (USTER Technologies Inc., Charlotte, NC) and an Advanced Fiber Information System (AFIS) (USTER Technologies Inc.) as described in Hinchliffe et al. (2010).

3. Results

3.1. Isolation of homeologous GhCesA2 genes from a G. hirsutum BAC library

For screening homeologous *GhCesA2* genes from a *G. hirsutum* TM-1 BAC library, a probe hybridizing the specific region of the *GhCesA2* cDNA was identified by searching *CesA* ESTs from the DFCI Cotton Gene Index (CGI, Release 10.1 at http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=cotton) containing 116,520 *Gossypium* genes. Among sixty-one *Gossypium* genes containing cellulose synthase signature motifs in the CGI database, two tentative consensus (TC) sequences (TC179659 and TC219062) were most similar to the 5′-truncated partial *GhCesA2* cDNA (U58284, 685 amino acids) that was an ortholog of *AtCesA4* (1,049 amino acids) and *PtiCes4* (1,042 amino acids) involved in SCW biosynthesis in *Arabidopsis* and *Populus*, respectively. By comparing all *Gossypium CesA*s in the CGI database, an overgo (708–747 nt of U58284) specifically binding to *CesA2* and a probe binding to a conserved *CesA* domain (1635–1800 nt of U58284) were used for screening *GhCesA2* homeologs. Through primary and secondary screenings, four BAC clones were identified as positive *GhCesA2* candidates from a BIBAC library for *G. hirsutum*, TM-1 covering 4.4x genome (Zhang et al., 2008). After fingerprinting and sequencing these four BAC clones, two different *GhCesA2* from allotetraploid *G. hirsutum* were identified and named *GhCesA2*-A_T (JN382209) and *GhCesA2*-D_T (JN382210) (Fig. 1). Computational analysis of the sequences from *GhCesA2*-A_T and *GhCesA2*-D_T showed that the putative transcriptional start sites of *GhCesA2*-A_T and *GhCesA2*-D_T were located at 216 and 222 nucleotides respectively upstream from the translational start codon, and the transcriptional start site was marked as +1 (Fig. 1). In both *GhCesA2*-A_T and *GhCesA2*-D_T, putative TATA boxes were located in the region −24/−29, and a putative CAAT box presented in the region −63/−66. Both *GhCesA2*-A_T and *GhCesA2*-D_T consisted of 12 exons and 11 introns (Fig. 1). They share 92% sequence similarity over the entire sequence. All splice sites from *GhCesA2*-A_T contained consensus splice dinucleotides (GT…AG), whereas one splice site of intron 10 from *GhCesA2*-D_T had nonconsensus GC splice donor dinucleotides instead of consensus GT donor dinucleotides (Fig. 1, *). The partial *GhCesA2* cDNA sequence (U58284; Pear et al., 1996) was identical to the coding sequences from exons 8–12 of *GhCesA2*-A_T, but different from those of *GhCesA2*-D_T.

To locate the two different *GhCesA2* genes in the subgenomes of allotetraploid cotton, *G. hirsutum*, specific primers for *GhCesA2*-A_T or *GhCesA2*-D_T were designed from the non-coding regions where sequence diversity occurred. The specific region of *GhCesA2*-A_T or *GhCesA2*-D_T was PCR amplified from various genomic DNA templates isolated from one allotetraploid *G. hirsutum* (AD_1), two A genome diploid cottons from *G. herbaceum* (A_1), *G. arborium* (A_2), and ten D genome diploid cottons from *G. thurberi* (D_1), *G. armourianum* ($D_{2\text{-}1}$), *G. davidsonii* ($D_{3\text{-}d}$), *G. klotzschianum* ($D_{3\text{-}k}$), *G. aridum* ($D_{4\text{-}3}$), *G. raimondii* (D_5), *G. gossypioides* ($D_{6\text{-}1}$), *G. lobatum* ($D_{7\text{-}4}$), *G. trilobum* ($D_{8\text{-}2}$), *G. laxum* ($D_{9\text{-}5}$) (Fig. 2). The specific region for *GhCesA2*-A_T was abundantly amplified from the templates of AD_1, A_1, and A_2 genomes, but rarely amplified from ten D genome species (Fig. 2). In contrast,

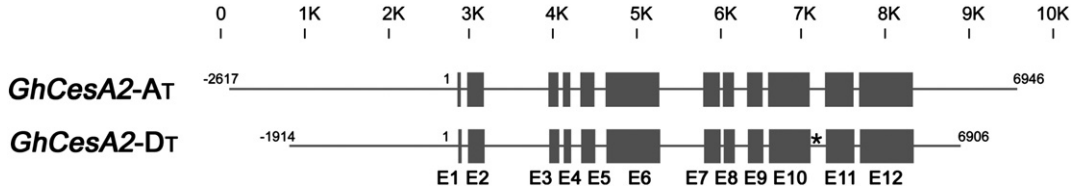

Fig. 1. Schemes of homeologous *GhCesA2* genes from allotetraploid *G. hirsutum*. The solid box represents exon sequences and the asterisk represents nonconsensus GC splice donor dinucleotides.

the specific region for *GhCesA2*-D_T was amplified from the templates of AD_1 and all ten D genomes, but was not amplified from either one of two A genome species (Fig. 2). The conserved *CesA* region was universally amplified from templates with AD_1, A, and D genomes (Fig. 2). In summary, we verified that *GhCesA2*-A_T and *GhCesA2*-D_T were homeologous genes located within the A-subgenome and D-subgenome, respectively.

3.2. Phylogenetic analysis of CesA genes from allotetraploid G. hirsutum

Phylogenetic relationships of *GhCesA2* homeologs from *G. hirsutum* were compared with *Arabidopsis thaliana CesA*s (*AtCesA*s) and *Populus trichocarpa CesA*s (*PtiCesA*s). *G. hirsutum* is a polyploid and perennial plant with long seed trichomes (fibers). *Arabidopsis* is a diploid and annual plant that lacks seed trichomes. In contrast, *P. trichocarpa* is a paleopolyploid and perennial tree containing seed trichomes. Due to its prolific production of seed trichomes, *P. trichocarpa* is also called "black cottonwood" (Tuskan et al., 2006). All 10 *AtCesA*s from the *Arabidopsis* genome were obtained from the TAIR site (http://www.arabidopsis.org/index.jsp). All 17 different *PtiCesA*s from the *P. trichocarpa* genome were downloaded from the JGI site (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). In addition to the *GhCesA2* homeologs, other published *GhCesA*s such as the *GhCesA1* homeologs (Pear et al., 1996; Kim and Triplett, 2001; Grover et al., 2004) and *GhCesA3* cDNA (Laosinchai et al., 2000; Zhu et al., 2011) clustered with *AtCesA*s and *PtiCesA*s.

Both *GhCesA1* homeologs and *GhCesA2* homeologs involved in SCW cellulose biosynthesis in developing cotton fibers clustered respectively with *AtCesA8* and *AtCesA4* involved in SCW cellulose biosynthesis in xylem from diploid *Arabidopsis* (Fig. 3). Interestingly, *GhCesA2*-A_T and *GhCesA2*-D_T from allotetraploid cotton were clustered with only one ortholog (*PtiCesA4*) from *Populus* although *GhCesA1*-A_T and *GhCesA1*-D_T clustered with two orthologs (*PitCesA8-A* and *PtiCesA8-B*) from *Populus*. *GhCesA3* aligned with one PCW *CesA* gene (*AtCesA3*) from diploid *Arabidopsis* and four PCW *CesA* genes (*PtiCesA3*s-A, *PtiCesA3*s-B, *PtiCesA3*s-C, and *PtiCesA3*s-D) from *Populus*.

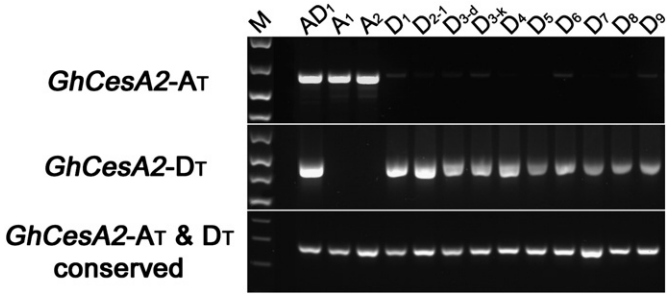

Fig. 2. Identification of homeologous *GhCesA2*-A_T and *GhCesA2*-D_T genes from allotetraploid *G. hirsutum*. The specific sequences for *GhCesA2*-A_T from the A-subgenome and *GhCesA2*-D_T from the D-subgenome of *G. hirsutum* were PCR amplified from following genomic DNAs: *G. hirsutum* TM-1 (AD_1), *G. herbaceum* (A_1), *G. arborium* (A_2), *G. thurberi* (D_1), *G. armourianum* ($D_{2\text{-}1}$), *G. davidsonii* ($D_{3\text{-}d}$), *G. klotzschianum* ($D_{3\text{-}k}$), *G. aridum* ($D_{4\text{-}3}$), *G. raimondii* (D_5), *G. gossypioides* ($D_{6\text{-}1}$), *G. lobatum* ($D_{7\text{-}4}$), *G. trilobum* ($D_{8\text{-}2}$), *G. laxum* ($D_{9\text{-}5}$). For the control (*GhCesA2*-A_T & D_T conserved), the conserved sequences among *CesA*s were amplified.

3.3. Temporal and spatial expression of GhCesA2-A_T and GhCesA2-D_T

The partial *GhCesA2* cDNA sequence (U58284), identical to the coding sequences between exon 8 and exon 12 from *GhCesA2*-A_T was previously reported to be expressed specifically in developing cotton fibers based on Northern analyses (Pear et al., 1996). With the availability now of more sensitive and specific methods of detecting transcript abundance, we tested if both *GhCesA2*-A_T and *GhCesA2*-D_T were specifically expressed in cotton fibers. We first compared the expression levels of *GhCesA2*-A_T and *GhCesA2*-D_T in various cotton tissues at different developmental stages using semi-quantitative RT-PCR with specific primer sets designed from the diverse 3′ UTR sequence of *GhCesA2*-A_T and *GhCesA2*-D_T (Fig. 4). Fig. 4A shows the specificity of the primer sets used in the semi-quantitative RT-PCR of *GhCesA2*-A_T and *GhCesA2*-D_T. The primer set for *GhCesA2*-A_T amplified the amplicon (424 bp) specifically from *GhCesA2*-A_T templates, whereas the primer set for *GhCesA2*-D_T amplified the amplicon (470 bp) specifically from *GhCesA2*-D_T template (Fig. 4A). Transcripts of *GhCesA2*-A_T and *GhCesA2*-D_T were commonly expressed in actively developing tissues like fiber (20 DPA), young leaves, 1-week old hypocotyls, 1-week old roots, bracts (DOA), and petals (DOA) (Fig. 4B). The highest level of *GhCesA2*-A_T and *GhCesA2*-D_T transcripts was detected in 20 DPA cotton fiber among the tested cotton tissues (Fig. 4B, Fiber). Ubiquitin transcripts that are expressed constitutively in various cotton tissues were used as a control.

Fig. 4C shows the temporal regulation of *GhCesA2*-A_T and *GhCesA2*-D_T during cotton fiber development. The expression levels of both *GhCesA2*-A_T and *GhCesA2*-D_T were rarely detected while rapid fiber elongation occurred (8–12 DPA), increased at the onset of secondary cell wall (SCW) cellulose biosynthesis (16 DPA), and remained at high levels during SCW biosynthesis (20 DPA) (Fig. 4C). In summary, both *GhCesA2*-A_T and *GhCesA2*-D_T shared common temporal and spatial expression patterns in *G. hirsutum*, and these genes were preferentially expressed during the SCW biosynthesis stage in developing cotton fibers.

3.4. Functional analysis of nonconsensus GC splice donor dinucleotides in GhCesA2-D_T

Several partial cotton *CesA* genes isolated for a comparative mapping project were defined as pseudogenes because they contained nonconsensus splice donor dinucleotides (Cronn et al., 1999). The nonconsensus splice dinucleotides were previously suggested to be involved in alternatively spliced genes in human and *C. elegans* (Thanaraj and Clark, 2001; Farrer et al., 2002). Thus, we questioned if alternative splice occurred in *GhCesA2*-D_T containing nonconsensus GC splice donor dinucleotides. Since alternative splice generates multiple types of transcripts from a single mRNA precursor, we tested first if any alternative *GhCesA2*-D_T mRNA isoforms were presented in developing cotton fibers. Using a specific primer set designed from 5′ and 3′ UTR of *GhCesA2*-D_T, the full length cDNA of *GhCesA2*-D_T was RT-PCR amplified from both elongation stage (8–12 DPA) and SCW cellulose biosynthesis stage (16–20 DPA) of cotton fibers. The full-length cDNA of *GhCesA2*-A_T was also amplified for comparing with *GhCesA2*-D_T since *GhCesA2*-A_T containing consensus splice donor dinucleotides alone should not have any alternative splices caused by

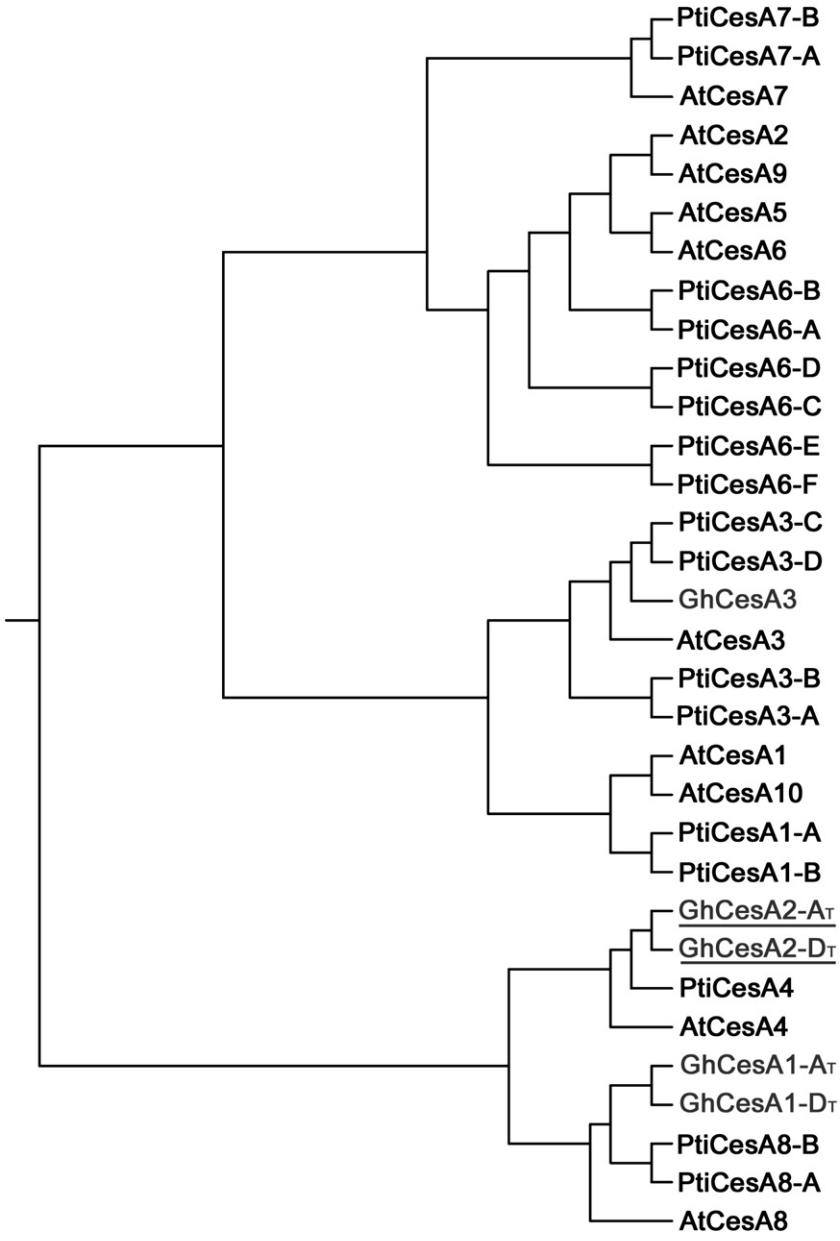

Fig. 3. Phylogenetic relationships of allotetraploid *G. hirsutum* cellulose synthase (*GhCesA*) genes and those from diploid *A. thaliana* (*AtCesA*) and paleopolyploid *P. trichocarpa* (*PtiCesA*). The phylogenetic tree was generated with the ClustalW program.

nonconsensus splice donor dinucleotides. Consistent with the semi-quantitative RT-PCR (Fig. 4C), full-length cDNAs of *GhCesA2*-A_T and *GhCesA2*-D_T were amplified specifically from the SCW cellulose biosynthesis stage (16 and 20 DPA) of cotton fibers (Fig. 5). The single amplicon band representing full length cDNA of *GhCesA2*-D_T (3,313 bp) as well as *GhCesA2*-A_T (3,264 bp) implied that the non-consensus GC splice donor dinucleotides of *GhCesA2*-D_T did not significantly contribute to alternative splice in developing cotton fibers when plants were grown on an irrigated field (Fig. 5). Both full length cDNAs were cloned and sequenced. The cDNA sequences of *GhCesA2*-A_T (JN382211) and *GhCesA2*-D_T (JN382212) showed that they have full open reading frames that were identical to all 12 exon sequences from *GhCesA2*-A_T gene (JN382209) and *GhCesA2*-D_T gene (JN382210) (Fig. 6). Thus, we concluded that *GhCesA2*-D_T containing nonconsensus GC splice donor dinucleotides was a functional gene as capable of generating functional *GhCesA2*-D_T transcripts as *GhCesA2*-A_T that contained consensus GT splice donor dinucleotides.

Deduced amino acid sequences between *GhCesA2*-A_T (1,039 amino acids) and *GhCesA2*-D_T (1,040 amino acids) have only four amino acids that are different (Fig. 6). The 5′-truncated partial *GhCesA2* cDNA sequence (U58284) encoding 685 amino acids, originally named *CelA2* (Pear et al., 1996) was identical to the *GhCesA2*-A_T cDNA sequence (Fig. 6). A protein motif search using InterProScan version 4.7 showed a zinc finger motif and eight transmembrane regions commonly found in *CesA*s (Fig. 6). Other distinct motifs (aspartate residues and Q/RXXRW) for CesA catalytic sites were also detected (Fig. 6).

3.5. Differential expression of GhCesA2-A_T and GhCesA2-D_T in cotton NILs showing different fiber bundle strength

Previously, our laboratory reported that there was differential gene expression in cotton fiber at the transition between elongation and SCW cellulose biosynthesis stages in two NILs having different fiber bundle strengths (Hinchliffe et al., 2010). Cotton line MD 52ne

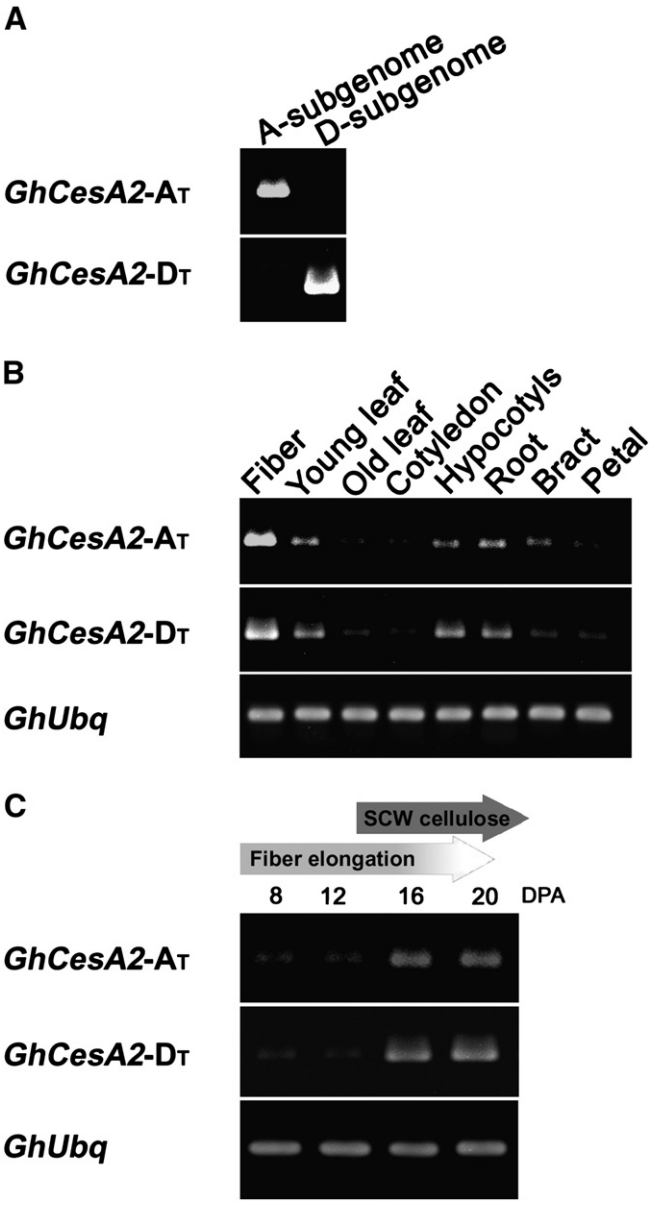

Fig. 4. Semi-quantitative RT-PCR of homeologous *GhCesA2* in allopolyploid *G. hirsutum*. (A) Primer specificity for *GhCesA2*-A_T and *GhCesA2*-D_T. The specific primer set for *GhCesA2*-A_T (amplicon, 424 bp) or *GhCesA2*-D_T (amplicon, 470 bp) was tested by PCR amplification with full length cDNA of *GhCesA2*-A_T (A-subgenome) or *GhCesA2*-D_T (D-subgenome) as a template. (B) Tissue preferential expression of *GhCesA2*-A_T and *GhCesA2*-D_T. The semi quantitative RT-PCR was performed with cDNA templates from young expanding leaves, old and fully expanded leaves, cotyledons, hypocotyls, root, bract, and petals. Ubiquitin transcripts expressed constitutively in various cotton tissues were used as a control. (C) Transcriptional regulation during fiber development. Semi quantitative RT-PCR was performed with cDNAs from different stages (8, 12, 16, and 20 DPA) of developing *G. hirsutum* fibers.

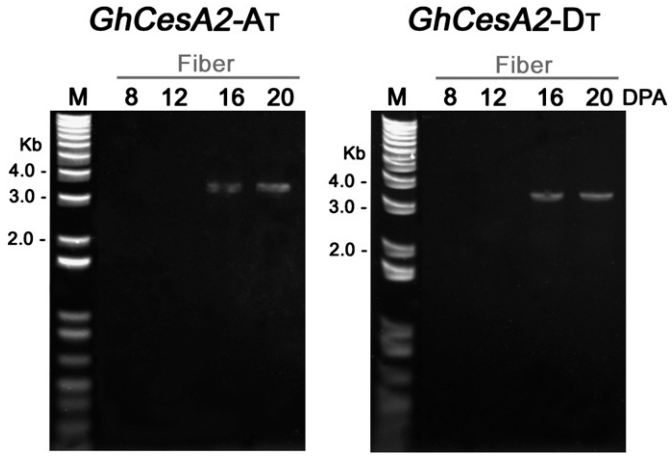

Fig. 5. Isolation of full length *GhCesA2*-A_T and *GhCesA2*-D_T cDNAs. Full length homeologous cDNAs of *GhCesA2*-A_T (3,264 bp) and *GhCesA2*-D_T (3,313 bp) were amplified by RT-PCR with a specific primer set and total RNAs isolated from 20 DPA fibers from *G. hirsutum*.

showed approximately 12% higher fiber-bundle strength than its NIL MD 90ne, but with no or little differences in average fiber length, fineness, and maturity (Meredith, 2005; Hinchliffe et al., 2010). In this study, we compared the expression levels of *GhCesA2*-A_T and *GhCesA2*-D_T between these two NILs by RT-qPCR (Fig. 7). Consistent with the results of semi-quantitative RT-PCR (Fig. 4C), the RT-qPCR results showed that transcript levels of both *GhCesA2*-A_T and *GhCesA2*-D_T were up-regulated at the onset of SCW cellulose biosynthesis in 14–16 DPA fibers of both MD 52ne and MD 90ne (Fig. 7A and B). Differential expression levels of *GhCesA2*-A_T and *GhCesA2*-D_T were detected between the two NILs. *GhCesA2*-A_T and *GhCesA2*-D_T transcript levels were higher in the higher bundle strength MD 52ne line at the transition between fiber elongation to secondary wall biosynthesis stages than for the lower bundle strength MD 90ne line. Fig. 7A shows that the difference in levels of *GhCesA2*-A_T transcripts in the higher bundle strength line (MD 52ne) were 2.0-fold at 14 DPA, 1.4-fold at 16 DPA, and 1.8-fold higher than those in the lower bundle strength line (MD 90ne). Almost identically, the levels of *GhCesA2*-D_T in the higher bundle strength line (MD 52ne) were 2.0-fold at 14 DPA, 1.4-fold at 16 DPA, and 1.9-fold higher than those in the lower bundle strength line (MD 90ne) (Fig. 7B).

4. Discussion

4.1. Identification of homeologous CesA2 from allotetraploid cotton

GhCesA2-A_T and *GhCesA2*-D_T were newly isolated from allotetraploid *G. hirsutum* (AD_1 genome). *GhCesA2*-A_T from the A-subgenome of *G. hirsutum* was similar to *CesA2* in A genome diploid species, whereas *GhCesA2*-D_T from the D-subgenome of *G. hirsutum* was similar to *CesA2* in D genome diploid species (Fig. 2). Comparison of *GhCesA2*-A_T and *GhCesA2*-D_T with partial *CesA2* genes from diploid *G. herbaceum* (A_1 genome) and diploid *G. raimondii* (D_5 genome) showed that sequences of *GhCesA2*-A_T and *GhCesA2*-D_T were conserved with *CesA2* from the A_1 genome and *CesA2* from the D_5 genome, respectively (Fig. S1). The open reading frame sequence between *GhCesA2*-A_T and *GhCesA2*-D_T was 99% identical (Fig. 6). *GhCesA2*-A_T and *GhCesA2*-D_T were clustered with *AtCesA4* and *PtiCesA4* whose expressions were abundantly detected in xylem cells containing SCW cellulose in *Arabidopsis* and *Populous*, respectively (Fig. 3). The 5′-truncated partial *GhCesA2* cDNA sequence (U58284), originally named *CelA2* (Pear et al., 1996), was identical to *GhCesA2*-A_T. In contrast, both *GhCesA2*-A_T and *GhCesA2*-D_T showed no significant sequence similarity with other partial *G. hirsutum cellulose synthase 2* genes (AF139448 and AF139450) that were used for evolutionary studies and comparative mapping and classified as *CesA2* pseudogenes due to the lack of consensus GT splice donor dinucleotides (Cronn et al., 1999; Senchina et al., 2003).

4.2. Potential roles of nonconsensus GC splice donor dinucleotides

Comparison of *GhCesA2* homeologs with partial *CesA2* genes from diploid *G. herbaceum* (A_1 genome) and diploid *G. raimondii* (D_5 genome) showed that the nonconsensus GC splice donor dinucleotides were conserved at intron 10 between *CesA2* from the diploid D_5 genome and *GhCesA2*-D_T from the allotetraploid D-subgenome

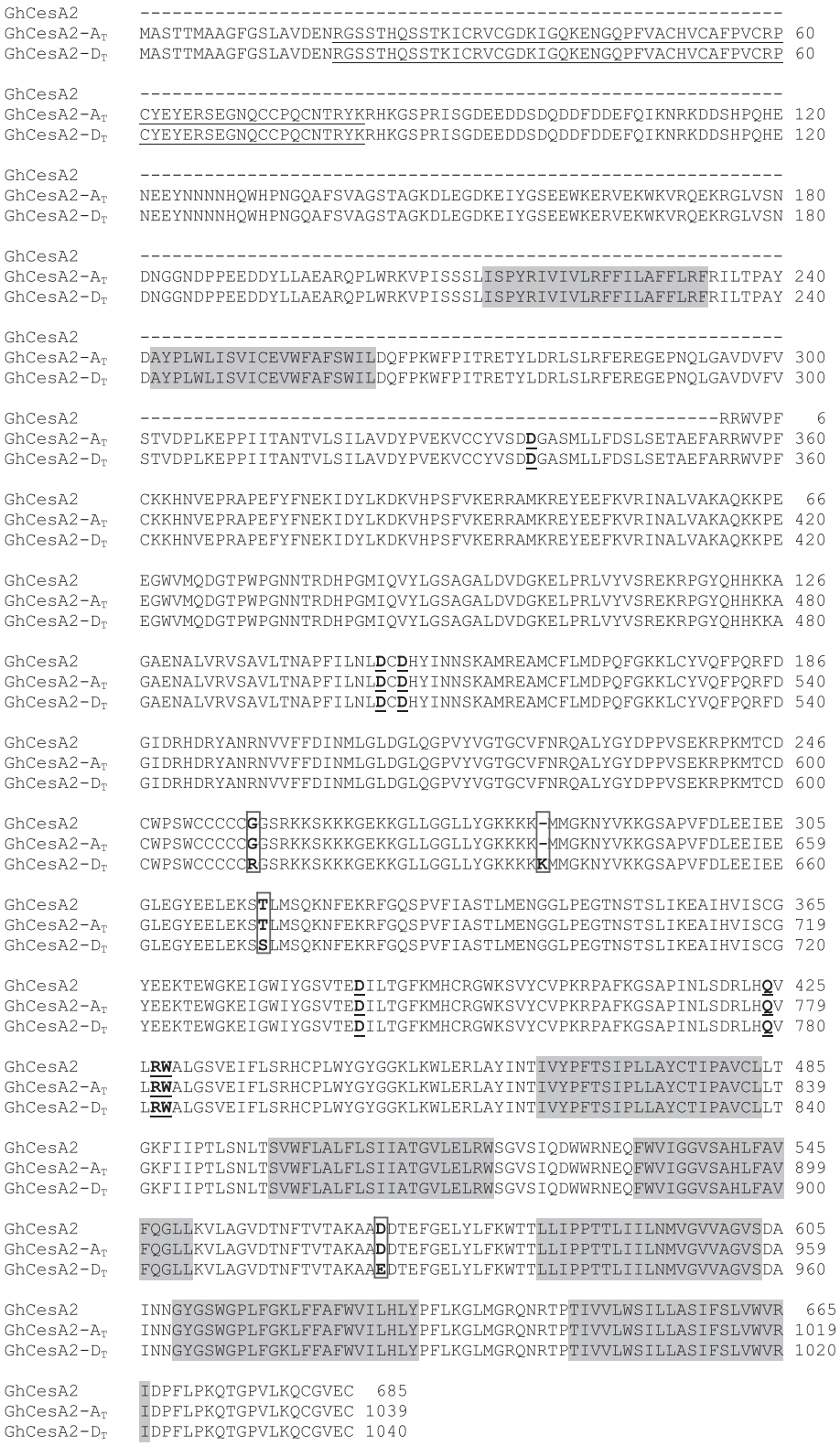

Fig. 6. Alignment of deduced amino acids from homeologous *GhCesA2* cDNAs of allotetraploid *G. hirsutum*. Different amino acids among the 5′-truncated partial GhCesA2 originally named as CelA2 (GhCesA2), *GhCesA2*-A_T, and *GhCesA2*-D_T were bold and boxed. A zinc finger motif was underlined, and eight transmembrane regions were highlighted. Catalytic sites of CesA (D and QXXRW) were bold and underlined.

(Fig. S1). In contrast, both *CesA2* genes from the diploid A_1 genome and *GhCesA2*-A_T from the allotetraploid A-subgenome contained consensus GT splice donor dinucleotides. In human genetics, aberrant splice due to mutations at consensus GT splice donor dinucleotides was suggested to cause human diseases from genetic disorders to cancer (Buratti et al., 2004; Krawczak et al., 2007; Roca et al., 2008). Nonconsensus splice dinucleotides have been suggested to be involved in alternative splice in humans, *C. elegans*, and *Arabidopsis* (Thanaraj and Clark, 2001; Farrer et al., 2002; Filichkin et al., 2010). Alternatively spliced mRNA isoforms harboring premature

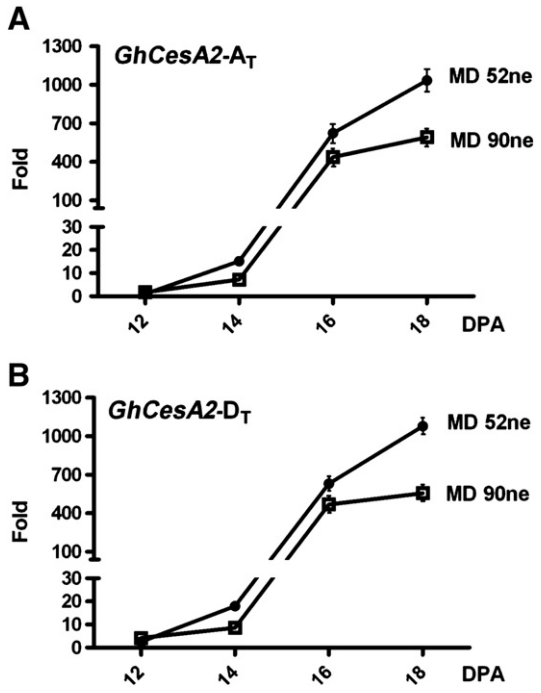

Fig. 7. Differential expression of homeologous *GhCesA2* in developing cotton fibers of near-isogenic lines MD52ne and MD90ne with different fiber bundle strengths. Transcript levels of *GhCesA2*-A_T (A) *and GhCesA2*-D_T (B) in developing cotton fibers (12–18 DPA) for *G. hirsutum*, MD 52ne and MD 90ne were compared by RT-qPCR. The transcript levels were normalized with respect to *18S ribosomal* RNA.

termination codons (PTCs) are selectively targeted and degraded by the nonsense-mediated mRNA decay (NMD) surveillance pathway (Chang et al., 2007). In polyploid genomes, one duplicated gene might degenerate to a pseudogene or acquire a modified function, so called neofunctionalization while the other duplicated gene kept its original function (Prince and Pickett, 2002; Vandepoele et al., 2003). Unlike the *GhCesA1* homeologs that clustered with two *CesAs* (*PtiCesA8-A* and *PtiCesA8-B*) among 17 *PtiCesAs* in the *Populous* genome, *GhCesA2* homeologs clustered with only one *Populus CesA* (*PtiCesA4*) (Fig. 3). It is possible that one *PtiCesA4* duplicated in an ancient genome might have degenerated after diploidization of the paleopolyploid *Populus* genome. Several cotton *CesA* genes from allotetraploid *G. hirsutum* (AF139448 and AF139450) as well as diploid *G. raimondii* (AF139449) and diploid *G. herbaceum* (AF139447) have been characterized as pseudogenes (Cronn et al., 1999) mainly due to the lack of consensus GT splice donor dinucleotides. In contrast, our results (Figs. 5 and 6) show that the *GhCesA2*-D_T gene containing nonconsensus splice donor dinucleotides was a functional gene that produced full-length *GhCesA2*-D_T mRNAs with little to no evidence for alternative mRNA isoforms in developing cotton fibers. In an attempt to find if alternatively spliced *GhCesA2*-D_T mRNA isoforms existed in the cotton EST database, we compared 41 ESTs encoding *CesA2* sequences identified from the database with *GhCesA2*-D_T. Sequence analyses showed that none of the identified ESTs were alternatively spliced from *GhCesA2* mRNA precursor. Among them, three ESTs (DW491531, DW491432, and CD486470) from *G. hirsutum* (AD_1 genome) and one EST (CO099632) from *G. raimondii* (D_5 genome) showed that the intron 10 containing the nonconsensus splice donor dinucleotides was successfully removed and two adjacent exons were properly ligated as shown by our results (Fig. S2). As a result, we conclude that *GhCesA2*-D_T containing nonconsensus splice donor dinucleotides was not a pseudogene. Although alternatively spliced *GhCesA2*-D_T mRNA isoforms were little detected in developing cotton fibers grown in irrigated field conditions, we do not rule out the possibility that the nonconsensus splice donor in *GhCesA2*-D_T may be involved in alternative splice in response to abiotic stresses such as drought and cold temperature in the same way that alternative splice of *Arabidopsis SRP30* is involved in responding to abiotic stresses (Filichkin et al., 2010). We are currently investigating if abiotic stresses reduce cellulose biosynthesis in developing cotton fibers by activating alternative splice processes of cotton *CesA* genes containing nonconsensus splice donor dinucleotides.

We also tested the possible existence of another copy of a *GhCesA2*-D_T gene in the *G. hirsutum* genome containing a consensus splice donor at intron 10 instead of a nonconsensus splice donor. The region covering intron 10 containing the nonconsensus splice donor dinucleotides of *GhCesA2*-D_T was PCR amplified from genomic DNA of allotetraploid *G. hirsutum* (AD_1). All of twenty-five *GhCesA2*-D_T amplified and cloned from *G. hirsutum* genomic DNA contained the nonconsensus GC splice donor dinucleotides. Since *GhCesA2*-D_T in the D-subgenome of allotetraploid *G. hirsutum* originated evolutionarily from *CesA2* in the D genome of diploid species *G. raimondii*, we also tested if there was another *CesA2* copy containing a consensus splice donor instead of a nonconsensus splice donor. All twenty-five *CesA2* genes amplified from *G. raimondii* contained the nonconsensus GC splice donor dinucleotides at intron 10. Thus, it is unlikely that another copy of the *GhCesA2*-D_T gene containing a consensus splice donor exists in *G. hirsutum*.

4.3. Transcriptional regulation of GhCesA2 homeologs

GhCesA2 originally named *Cel2* was previously reported to be a cotton fiber specific gene by Northern blot analysis (Pear et al., 1996). Semi-quantitative RT-PCR (Fig. 4) and RT-qPCR (Fig. 7) that are more sensitive in detecting low abundant transcripts than Northern blot analysis showed that *GhCesA2* was not specifically, but rather preferentially expressed in cotton fibers (Fig. 4). These results were consistent with the expression patterns of *GhCesA4* from the A-subgenome of *G. hirsutum* being a homeologous gene of *GhCesA1* from the D-subgenome of *G. hirsutum* (Pear et al., 1996; Kim et al., 2002; Grover et al., 2004). Northern blot analysis showed that *GhCesA4* was expressed specifically in cotton fibers (Kim et al., 2002), but RT-qPCR and promoter analyses with a reporter gene showed *GhCesA4* was expressed in both fiber and non-fiber tissues although *GhCesA4* transcripts were highly expressed during the SCW biosynthesis stage of fiber development (Kim et al., 2011). The results of Figs. 4, 5, and 7 showed that transcript levels of *GhCesA2* homeologs were commonly regulated in different developmental stages from various cotton tissues.

The higher expression (1.4–2.0 fold) of both *GhCesA2*-A_T and *GhCesA2*-D_T in the higher bundle strength MD 52ne line than the lower bundle strength MD 90ne line at the transition stage (14–18 DPA) was consistent with the previous microarray results showing that *GhCesA2* transcript levels in the MD52ne line at the transition stage (16 DPA) were 2.1-fold higher than those in the MD90ne line (Fig. 7; Hinchliffe et al., 2010). The microarray analyses between two NILs revealed that SCW biosynthesis-related genes were significantly up-regulated at the transition stage (16 DPA) of cotton fiber in the higher bundle strength MD52ne line over the lower bundle strength MD 90ne (Hinchliffe et al., 2010). The higher transcript levels of SCW biosynthesis-related genes in developing fibers from MD52ne were suggested to contribute higher fiber strength by affecting the degree of polymerization of cellulose or/and cellulose microfibril orientation (Hinchliffe et al., 2010). Both *GhCesA2*-A_T and *GhCesA2*-D_T up-regulating at the transition stage of developing cotton fibers are mainly responsible for SCW biosynthesis in cotton fibers (Pear et al., 1996). In *Arabidopsis*, a mutation of *AtCesA4* that is an ortholog of *GhCesA2*-A_T and *GhCesA2*-D_T resulted in a SCW deficient phenotype (*irregular xylem 5*) in which the cellulose content was reduced to 30% of wild type *Arabidopsis* (Taylor et al., 2003). Thus,

differential expression of *GhCesA2*-A_T and *GhCesA2*-D_T may affect SCW biosynthesis of cotton fiber.

In summary, we isolated two *CesA2* homeologous genes, *GhCesA2*-A_T from the A-subgenome and *GhCesA2*-D_T from the D-subgenome of allotetraploid *G. hirsutum*. Both *GhCesA2*-A_T and *GhCesA2*-D_T could produce functional and full-length transcripts despite the existence of nonconsensus splice donor dinucleotides in *GhCesA2*-D_T. *GhCesA2*-A_T and *GhCesA2*-D_T were preferentially and most abundantly expressed during the cellulose biosynthesis stage of developing cotton fiber and might be involved in regulating cotton fiber properties. Thus, the ability to manipulate the expression of *GhCesA*s involved in SCW cellulose biosynthesis may lead to strategies for improving cotton fiber quality.

Supplementary materials related to this article can be found online at doi:10.1016/j.gene.2011.12.018.

Acknowledgments

This work was supported by the USDA-ARS, the National Science Foundation, and the Louisiana State Support Program of Cotton Incorporated. The authors gratefully acknowledge Drs. Richard Percy and Lori Hinze from USDA-ARS-SPARC for providing genomic DNAs of diploid cotton species and Ms. Fanny Liu from USDA-ARS-MSA for sequencing BAC clones. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

References

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U. S. A. 100, 4649–4654.

Applequist, W.L., Cronn, R., Wendel, J.F., 2001. Comparative development of fiber in wild and cultivated cotton. Evol. Dev. 3, 3–17.

Buratti, E., Chivers, M., Kralovicova, J., Romano, M., Baralle, M., Krainer, A.R., Vorechovsky, I., 2004. Aberrant 5′ splice sites in human disease genes: Mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Res. 35, 4250–4263.

Burn, J.E., Hocart, C.H., Birch, R.J., Cork, A.C., Williamson, R.E., 2002. Functional analysis of the cellulose synthase genes *CesA1*, *CesA2*, and *CesA3* in *Arabidopsis*. Plant Physiol. 129, 797–807.

Burton, R.A., Shirley, N.J., King, B.J., Harvey, A.J., Fincher, G.B., 2004. The CesA gene family of barley quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134, 224–236.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622.

Chang, Y.-F., Imam, J.S., Wilkinson, M.F., 2007. The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. 76, 51–74.

Chaudhary, B., Flagel, L., Stupar, R.M., Udall, J.A., Verma, N., Springer, N.M., Wendel, J.F., 2009. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182, 503–517.

Cronn, R.C., Small, R.L., Wendel, J.F., 1999. Duplicated genes evolve independently following polyploid formation in cotton. Proc. Natl. Acad. Sci. U. S. A. 96, 14406–14411.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. 104, 482–489.

Doblin, M., Kurek, I., Jacob-Wilk, D., Delmer, D., 2002. Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol. 43, 1407–1420.

Farrer, T., Roller, A.B., Kent, W.J., Zahler, A.M., 2002. Analysis of the role of *Caenorhabditis elegans* GC-AG introns in regulated splicing. Nucleic Acids Res. 30, 3360–3367.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.-K., Mockler, T.C., 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res. 20, 45–58.

Grover, C.E., Kim, H., Wing, R.A., Paterson, A.H., Wendel, J.F., 2004. Incongruent patterns of local and global genome size evolution in cotton. Genome Res. 14, 1474–1482.

Hinchliffe, D.J., Meredith, W.R., Yeater, K.M., Kim, H.J., Woodward, A.W., Chen, Z.J., Triplett, B.A., 2010. Near-isogenic cotton germplasm lines that differ in fiber-bundle strength have temporal differences in fiber gene expression patterns as revealed by comparative high-throughput profiling. Theor. Appl. Genet. 120, 1347–1366.

Hovav, R., Udall, J.A., Chaudhary, B., Rapp, R., Flagel, L., Wendel, J.F., 2008. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc. Natl. Acad. Sci. U. S. A. 105, 6191–6195.

Kim, H.J., Triplett, B.A., 2001. Cotton fiber growth *in planta* and *in vitro*: models for plant cell elongation and cell wall biogenesis. Plant Physiol. 127, 1361–1366.

Kim, H.J., Williams, M.Y., Triplett, B.A., 2002. A novel expression assay system for fiber specific promoters in developing cotton fibers. Plant Mol. Biol. Rep. 20, 7–18.

Kim, H.J., Murai, N., Fang, D.D., Triplett, B.A., 2011. Functional analysis of *Gossypium hirsutum* cellulose synthase catalytic subunit 4 promoter in transgenic *Arabidopsis* and cotton tissues. Plant Sci. 180, 323–332.

Krawczak, M., Thomas, N.S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., Cooper, D.N., 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum. Mutat. 28, 150–158.

Kumar, M., Thammannagowda, S., Bulone, V., Chiang, V., Han, K.-H., Joshi, C.P., Mansfield, S.D., Mellerowicz, E., Sundberg, B., Teeri, T., Ellis, B.E., 2009. An update on the nomenclature for the cellulose synthase genes in *Populus*. Trends Plant Sci. 14, 1360–1385.

Laosinchai, W., Cui, X., Brown Jr., R.M., 2000. A full length cDNA of cotton cellulose synthase has high homology with the Arabidopsis RSW1 gene and the cotton CelA1 gene (Accession No. AF200453) (PGR 00–002) Plant Physiol. 122, 291.

Meinert, M., Delmer, D.P., 1977. Changes in biochemical composition of the cell wall of the cotton fiber during development. Plant Physiol. 59, 1088–1097.

Meredith, W.R., 2005. Registration of MD 52ne high fiber quality cotton germplasm and recurrent parent MD 90ne. Crop. Sci. 45, 806–807.

Mueller, O., Lightfoot, S., Schroeder, A., 2004. RNA integrity number (RIN)—standardization of RNA quality control. Agilent Application Note, Publication 5989-1165EN, pp. 1–8.

Naithani, S.C., Rama-Rao, R.N., Singh, Y.D., 1982. Physiological and biochemical changes associated with cotton fibre development I. Growth kinetics and auxin content. Physiol. Plant. 54, 225–229.

Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P., Stalker, D.M., 1996. Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. U. S. A. 93, 12637–12642.

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M., Somerville, C.R., 2007. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 104, 15566–15571.

Prince, V.E., Pickett, F.B., 2002. Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet. 3, 827–837.

Richmond, T.A., Somerville, C.R., 2000. The cellulose synthase superfamily. Plant Physiol. 124, 495–498.

Roca, X., Olson, A.J., Rao, A.R., Enerly, E., Kristensen, V.N., Børresen-Dale, A.-L., Andresen, B.S., Krainer, A.R., Sachidanandam, R., 2008. Features of 5′-splice-site efficiency derived from disease-causing mutations and comparative genomics. Genome Res. 18, 77–87.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A., Wendel, J.F., 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20, 633–643.

Soderlund, C., Longden, I., Mott, R., 1997. FPC: A system for building contigs from restriction fingerprinted clones. CABIOS 13, 523–535.

Somerville, C., 2006. Cellulose synthesis in higher plants. Annu. Rev. Cell Dev. Biol. 22, 53–78.

Tanaka, K., Murata, K., Yamazaki, M., Onosato, K., Miyao, A., Hirochika, H., 2003. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol. 133, 73–83.

Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K., Turner, S.R., 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. U. S. A. 100, 1450–1455.

Thanaraj, T.A., Clark, F., 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res. 29, 2581–2593.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., et al., 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313, 1596–1604.

Vandepoele, K., Simillion, C., Van, D.P.Y., 2003. Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15, 2192–2202.

Wendel, J.F., Cronn, R.C., 2003. Polyploidy and the evolutionary history of cotton. Adv. Agron. 78, 139–186.

Zhang, H.-B., Li, Y., Wang, B., Chee, P.W., 2008. Recent Advances in Cotton Genomics. Int. J. Plant Genomics 2008, 742304.

Zhu, H., Han, X., Lv, J., Zhao, L., Xu, X., Zhang, T., Guo, W., 2011. Structure, expression differentiation and evolution of duplicated fiber developmental genes in *Gossypium barbadense* and *G. hirsutum*. BMC Plant Biol. 11, 40.